Year: 2018

Removal of irrelevant information from working memory: sometimes fast, sometimes slow, and sometimes not at all

Oberauer, Klaus

Abstract: To function properly, working memory must be rapidly updated. Updating requires the removal of information no longer relevant. I present six experiments designed to explore the boundary conditions and the time course of removal. A condition in which three out of six memory items can be removed was compared to two baseline conditions in which either three or six items were encoded and maintained in working memory. The time for removal was varied. In experiment 1, in the removal condition, a distinct subset of three words was cued to be irrelevant after encoding all six words. With longer removal time, response times in the removal condition approximated those in the set-size 3 baseline, but accuracies stayed at the set-size 6 level. In experiment 2, in which a random subset of three words was cued as irrelevant, there was no evidence for removal. Experiments 3 and 4 showed that when each item is cued as relevant or irrelevant after its encoding, irrelevant items can be removed rapidly and completely. Experiments 5 and 6 showed that complete removal was no longer possible when words had to be processed before being cued as irrelevant. The pattern of findings can be explained by distinguishing two forms of removal: deactivation removes working-memory contents from the set of competitors for retrieval; unbinding contents from their contexts removes them from working memory entirely, so that they also cease to compete for limited capacity.

DOI: https://doi.org/10.1111/nyas.13603

Originally published at:
Oberauer, Klaus (2018). Removal of irrelevant information from working memory: sometimes fast, sometimes slow, and sometimes not at all. Annals of the New York Academy of Sciences, 1424(1):239-255.
DOI: https://doi.org/10.1111/nyas.13603

Removal of Irrelevant Information from Working Memory:

Sometimes Fast, Sometimes Slow, and Sometimes Not at All

Klaus Oberauer

University of Zurich

Oberauer, K. (2018). Removal of irrelevant information from working memory: sometimes fast, sometimes slow, and sometimes not at all. *Annals of the New York Academy of Science, 1424*, 239–255. doi:10.1111/nyas.13603

Author Note

Klaus Oberauer, Department of Psychology, University of Zurich, Switzerland. This research was supported by a grant from the Swiss National Science Foundation to K. Oberauer (project 149193). All data are available on the Open Science Framework (OSF) at osf.io/vehmc. Correspondence should be addressed to Klaus Oberauer, Department of Psychology, Cognitive Psychology Unit, University of Zürich, Binzmühlestrasse 14/22, 8050 Zurich, Switzerland. E-mail: k.oberauer@psychologie.uzh.ch

Abstract

To function properly, working memory must be rapidly updated. Updating requires removal of information no longer relevant. I present six experiments designed to explore the boundary conditions and the time course of removal. A condition in which 3 out of 6 memory items can be removed was compared to two baseline conditions in which either 3 or 6 items were encoded and maintained in working memory. The time for removal was varied. In Experiment 1, in the removal condition a distinct subset of 3 words was cued to be irrelevant after encoding all six words. With longer removal time, RTs in the removal condition approximated those in the set-size 3 baseline, but accuracies stayed at the set-size 6 level. In Experiment 2, in which a random subset of 3 words was cued as irrelevant, there was no evidence for removal. Experiments 3 and 4 showed that when each item is cued as relevant or irrelevant after its encoding, irrelevant items can be removed rapidly and completely. Experiments 5 and 6 showed that complete removal was no longer possible when words had to be processed before being cued as irrelevant. The pattern of findings can be explained by distinguishing two forms of removal: De-activation removes working-memory contents from the set of competitors for retrieval; unbinding contents from their contexts removes them from working memory entirely, so that they also cease to compete for limited capacity.

Keywords: working memory; removal; forgetting; binding;

Removal of Irrelevant Information from Working Memory: Sometimes Fast, Sometimes Slow, and Sometimes Not at All

Working memory (WM) is the system that holds mental representations available for processing. Because WM has a limited capacity [1], to work well, it must limit itself to the information most relevant for ongoing cognitive activity at any moment in time [2]. Our thoughts and actions progress rapidly – for instance, when reading, we take in about one new concept every 300 to 500 ms [3], and when exploring a visual scene, we fixate new objects at about the same rate [4]. Therefore, the information most relevant changes rapidly. WM must update its contents efficiently, and this includes getting rid of old, no-longer relevant contents quickly lest it gets clogged up with outdated information [5, 6].

Until recently the process of clearing WM from outdated information has received little attention, probably because the dominant view was that representations in WM decay rapidly by default. Some have argued explicitly that decay serves the function of "garbage collection" in WM [7]. To date, the balance of evidence speaks strongly against any role for decay in WM for verbal materials [8-11]. The evidence is more favorable to decay for visual and spatial information, for which several experiments observed a gradual decline over an unfilled retention interval [12-15]. If we take this decline to reflect decay, and estimate the decay rate from the rate of decline of performance, it is too slow to serve the function of clearing out irrelevant information. For instance, accuracy in recognizing visual stimuli declined from about 85% to about 75% over 12 s in Ricker, Spiegel [13]. In free scene perception we would fixate – and by implication, attend to – more than 30 new objects during that time. If WM is to maintain the most recently attended objects, and got rid of old objects through decay at a rate as indicated in these experiments, it would maintain reasonably strong representations of more than 30 objects at any moment in time – an order of magnitude more than common estimates for the capacity of WM for visual objects.

These considerations imply that efficient updating of WM requires a process of rapid removal of outdated information. Removal differs from decay in that it targets selectively representations no longer needed for the current task. Hasher, Zacks, and colleagues have proposed a similar process – deletion of outdated information from WM – as one of three forms of inhibitory processes that are impaired in old age, contributing to poorer WM functioning in older adults [6]; there is some evidence this deletion process is impaired in old age [16]. Recent research with a WM updating paradigm has identified selective removal as a component process of updating [17-20].

Other research has begun to explore removal with a retro-cue paradigm: Participants encode a set of items into WM, and subsequently receive a cue identifying one item, or a subset of items, as relevant; by implication, the remaining contents of WM are irrelevant and can be removed. In one retro-cue paradigm, participants initially encode two subsets of items into WM, and are then cued to remember one of them, selected at random. After a variable cue-target interval, a recognition probe is presented, and participants are to decide whether it matches an item in the relevant subset (rejecting probes matching an item in the irrelevant set). After a cue-target interval of 1s or more, the size of the irrelevant subset ceases to affect response times to the probe, suggesting that the irrelevant subset was removed during that interval [21]. Another retro-cue paradigm starts with encoding a single homogeneous set of items into WM, of which a single item is subsequently cued as relevant, allowing people to remove the remaining items [22, 23]. Evidence for removal comes from the finding that, after a retro-cue, people find it easier to add further information to WM [24]. Moreover, when, on rare occasions, memory for an item previously cued as irrelevant is tested, it is extremely poor [25, 26].

Removal of irrelevant information from WM also plays an important role in a computational model of WM, the SOB-CS model [27]. In that model, memory lists are encoded into WM by rapidly binding each item to a context representing its serial position in the list. Some experimental paradigms – for instance, the complex-span paradigm [28, 29] – require participants to process additional materials during the retention interval (e.g., reading sentences, solving arithmetic equations). In SOB-CS, representations involved in these processing tasks – called *distractors* – are initially encoded into WM in the same way as memory items, by binding them to the current position context. Because distractors are no longer needed after they have been processed, they can subsequently be removed through unbinding them from their context. This removal is assumed to be gradual, taking about 1-2s to completion. Evidence for these assumptions comes from experiments showing that memory items tend to be confused with distractors at recall, and this tendency for distractors to intrude is reduced when more time is allowed for removing them from WM after processing [30].

The present series of experiments aimed to delineate the boundary conditions of selective removal of information from WM, and to measure the speed and effectiveness of removal. To this end I used a variant of the two-set retro-cue paradigm [21]. Participants initially encoded six words into WM; each word was presented in one of six frames on the screen (see Figure 1). Subsequently, a subset of three words was cued as relevant (to-be-remembered), implying that the other three words can be forgotten. After a variable time following the cue (cue-target interval, CTI), memory for one randomly selected word from the relevant set was tested with a local-recognition probe: The probe word was presented in one of the relevant frames, and participants decided whether it matched the word previously presented in that frame. The removal condition, in which the initial memory set size of six words could be reduced to three by removing the other three words, (referred to as *set size 6-3*), was compared to two control conditions: One in which all six words are cued as relevant (set size 6), and one in which only three words were presented, all of which remain relevant (set size 3). By comparing performance (accuracy and response times) in the set-size 6-3 condition to the two control conditions I can measure how effectively the irrelevant subset of words was removed at any given CTI: To the extent that the irrelevant three words have been removed from WM in the set-size 6-3 condition, performance should become better than in the set-size 6 condition and approach that in the set-size 3 condition.

The raw data from all experiments are available on the Open Science Framework (OSF) at osf.io/vehmc.

Experiment 1

The first experiment was very similar to the original two-set paradigm [21]. Participants encoded two sets of three words, displayed across two rows of frames, one above the other. One subset was cued as relevant.

Method

Participants. Twenty-four students of the University of Zurich took part in two one-hour sessions in exchange for course credit or 30 Swiss Francs (approx. 30 US $).

Materials. For each trial six words were drawn at random, without replacement within a session, from a set of German words, 4-5 letters long.

Procedure. Each trial started with the presentation of two rows of three rectangular frames each, one in the upper and one in the lower half of the screen. The frames were displayed with thin black outlines on a white background (see Figure 1, left side). Presentation of the memory words commenced 1s later. The six words were presented one by one, for 1s each, from left to right across the three frames of the top row, followed by the bottom row. An additional 0.5 s pause, during which all frames remained empty, was inserted between the first and the second subset of words to enhance their separation into two groups [31]. In the set-size 3 condition, one of the subsets, selected at random with equal probability, was replaced by three copies of "XXXX", displayed in the same way as the words. One second after the offset of the last word (or "XXXX"), the frames of one row (set-size 3 and set-size 6-3), or of both rows (set-size 6), turned red, indicating the relevant (to-be-remembered) set. In the set-size 3 condition the relevant row was always the one with words; in the set-size 6-3 condition, the relevant row was selected at random, with an equal chance, on each trial. Participants were instructed that only words from the relevant set will be tested, and that therefore they could "forget" the not-cued subset.

After a variable cue-target interval (CTI), a recognition probe was presented in a randomly selected frame within the relevant set. On half the trials this word matched the word originally presented in that frame; on the other half it matched a word in one of the other frames of the relevant set. Thus, participants had to remember not only which words they have seen, but also in which frame each word had been presented. In this way, their memory could not rely on familiarity of the words but rather had to rely on recollection of the word-location bindings. I believe that this is important for measuring the load on WM capacity because individual differences in WM capacity, measured through a separate set of tasks, correlates more strongly with recollection than with familiarity in the local-recognition task [32]. Participants entered their response by pressing the left arrow key for "mismatch", and the right arrow key for "match". They were instructed to respond as quickly as possible without committing unnecessary errors.

The CTI was varied over five levels: 0.1, 0.4, 0.8, 1.5, and 3.0s. CTI was varied between blocks, so that within a block, participants could anticipate how much time they will have between the cue and the probe. In this way I hoped to provide them with optimal conditions for using that time to remove the irrelevant set from WM. In each session participants worked through five blocks of 36 trials each. The order of blocks was counterbalanced across participants using a Latin square. Each block consisted of 12 trials of each set-size condition, half of which were tested with a positive (matching) and the other half with a negative (not-matching) probe. Within each block, the test trials were preceded by 6 practice trials (2 from each condition) with that block's CTI.

Data Analysis

I assessed memory performance through both response time (RT) and accuracy (proportion correct). For RTs, I first removed all error trials and all RTs exceeding 10s. In a second step I removed all RTs exceeding a person's mean in each set-size condition per session by 3 intra-individual SDs. Finally, I log-transformed the remaining RTs to obtain a more normal distribution. Each dependent variable was analyzed through two Bayesian Analyses of Variance (BANOVAs) with condition and CTI as independent variables. One BANOVA compared the set-size condition 6-3 to the set-size 6 baseline, and the other compared it to the set-size 3 baseline. The BANOVAs used the *BayesRS* package [33] for *R* [34]. The BANOVA model in the *BayesRS* package largely follows Rouder, Morey

[35], but includes random slopes (i.e., individual differences in the size of the main effects and interactions) for categorical and continuous predictors. This extension was necessary to enter CTI as a continuous linear predictor, and allowing for random slopes of that predictor. The BANOVA model used Cauchy default priors with scale 1/2 for the effect sizes of categorical predictors, and of sqrt(2)/4 for the effect size of the continuous predictor.

I computed the Bayes Factor (BF) in favor of each main effect and the interaction through the Savage-Dickey approximation [36]. The BF reflects the strength of evidence in the data for the effect in question; it is the factor by which the prior odds for a model including the effect, relative to an otherwise identical model excluding it, should be updated to obtain their posterior odds. The inverse of the BF in favor of an effect is the strength of evidence against the effect (i.e., for excluding the effect from the model). The Savage-Dickey approximation yields the BF in favor of an effect as the ratio of the posterior to the prior of the effect size in question, evaluated at 0. The density of the posterior was obtained from a normal distribution fitted to 50,000 MCMC samples of the posterior.

Results

Mean RT and proportion correct are presented in Figure 2. Table 1 provides the Bayes Factors in favor of the main effects of the two set-size contrasts of interest (6-3 vs. 6, and 6-3 vs. 3), of CTI, and the interaction of CTI with each of the set-size contrast. BFs *against* each effect can be obtained by taking the reciprocal of the BFs in the table.

The RT data of the 6-3 condition show the trajectory expected from gradual, and eventually complete, removal of the irrelevant set: They start at the level of set size 6 at the shortest CTI, and reach the level of set size 3 at a CTI of 1.2s. This finding converges with my earlier observation with the two-list paradigm using a global-recognition test of memory (i.e., the "modified Sternberg" paradigm), in which the set-size effect of the irrelevant list on RTs declined to zero within 1 to 1.5s [21]. The accuracy data tell a different story: Accuracy in the set-size 6-3 condition did not depart credibly from those in the set-size 6 condition. I did not analyze accuracies in the 2001 study because accuracies were too close to ceiling, probably because global recognition is easier than local recognition. The accuracy data are in line with earlier work on directed forgetting of short lists of nonword-word pairs: After encoding two subsets of pairs, distinguished by their color, cueing one of them as to-be-forgotten resulted in only a very small improvement of accuracy for the remaining subset [5].

Discussion

The results of Experiment 1 are consistent with the findings from my earlier study [21] and its subsequent replication [37], but the new findings on accuracies requires a qualification of my earlier interpretation: The irrelevant sub-set is not removed from WM in such a way that it undoes encoding.

I suggest the following tentative explanation for the present pattern of results [see also 38]: The set-size effect on RTs reflects the competition between several representations in WM for being accessed and compared to the probe. Resolving that competition takes time, but as long as it is always resolved in favor of the correct item, it does not affect accuracy. The set-size effect on accuracy reflects the capacity limit on WM. The capacity limit arises from interference between items maintained simultaneously in WM – more precisely, between the bindings of each item to its context [27, 39]. The term *context* refers to the information that is used to selectively accessing an item in WM; in the present paradigm that could be the word's serial position of presentation, or its spatial location on the screen, because the two are perfectly correlated. Interference results in a distortion of the representations retrieved from WM. Comparing the probe to a more distorted representation of the original word in the probed frame results in more errors. If these assumptions are correct, the present findings indicate that the irrelevant set is removed from the set of items competing for retrieval, but not from the set of items that interfere with each other in WM.

How could the set of items competing for retrieval be a proper subset of the set of items interfering with each other in WM? Consider the representational structure built in WM when encoding a set of six items, organized into two groups (Figure 3).[1] Items are bound to two kinds of context representations on different levels of a hierarchy of contexts [40, 41]: On a first level, each item is bound to a context representing its position (1 to 6) in the memory set. On a second level, each item is bound to a context representing its group (top or bottom). After encoding, both group contexts remain moderately active, thereby activating all six items. When a retro-cue indicates only the top group as relevant, activation of the bottom context subsides, and only the three items bound to the top context remain active. At test, the frame in which the probe is presented identifies a single position as relevant, so that the corresponding position context is now strongly activated, thereby giving the target item an activation boost, by which it wins the retrieval competition.

In this scenario, the set of active item representations constitutes the set of items competing for retrieval. The group cue enables limiting this set to the items bound to one group context – the three items in the other group are so much disadvantaged by their low level of activation that they effectively drop out of the competition. At the same time, all six items interfere with each other to the extent that their contexts overlap [27]. This interference occurs between the bindings of items to their contexts: All these bindings are superimposed in a common representational medium (i.e., a matrix of connection weights in a neural network), and thereby distort each other. I assume that a group cue presented after encoding of all list items does only help to reduce the current activation of items bound to the irrelevant group, but does not enable removing the bindings of these items to their context. Therefore, the group cue reduces the competition for retrieval, but not the degree of interference between item-context bindings.

The assumption that a group cue presented after encoding of all memory items is not suited to remove the bindings of the irrelevant items from their context is motivated by the computational mechanism implementing removal of item-context bindings in the SOB-CS model [27]. In this model, bindings are established through rapid Hebbian learning, and they can be removed through Hebbian

[1] Figure 3 is a simplified sketch of the memory representation in SOB-CS, with the additional assumption of group contexts. Whereas the figure depicts each representation as a separate unit, SOB-CS uses overlapping distributed representations.

anti-learning (i.e., Hebbian learning with a negative learning rate). Both Hebbian learning and anti-learning require that the two representations to be bound to each other are currently active. During presentation of each memory item, that item and its context are active (and no other item or context is), so that they can be bound to each other through Hebbian learning. As long as that item and its context remain active, their binding could also be removed through Hebbian anti-learning (effectively, continuing the learning process after flipping the sign of the learning rate). However, at a later point in time, when all items have been encoded, going back to a previously encoded item to remove it from its context would require to selectively re-activate that item and its context, and then apply Hebbian anti-learning to that pair. This could be done by retrieving the to-be-removed items one by one. Doing so, however, is risky because retrieval can fail, and in that case, the wrong item's bindings would be removed, potentially damaging the representation of the relevant group. Therefore, I assume that, by default, the WM system is poised to remove the item-context binding of a stimulus immediately after encoding, but refrains from trying to remove item-context bindings for any previously encoded stimulus.

The theoretical ideas outlined above entail two predictions. First, a group cue should enable limiting the set of activated items to the relevant subset if and only if the items in that subset have been bound to a group context, so that the irrelevant group's context can be selectively switched off. Therefore, a cue to a subset defined *ad hoc* at the time of cueing should not enable reducing the set of retrieval competitors. Experiment 2 will test that prediction. Second, removal of item-context bindings should be much easier if the cue indicating whether or not an item should be remembered is presented immediately after encoding each item, rather than after encoding the entire memory set. This prediction will be tested with Experiments 3 and 4.

Experiment 2

In Experiment 1, the memory set consisted of two clearly distinguishable groups at encoding. This facilitates the formation of group contexts to which the items of each group are bound. In Experiment 2, the memory set is presented as a homogenous set of six words, and when (in the 6-3 condition) a subset is identified as to-be-remembered, it is a random subset on each trial. Therefore, the WM system has no chance to bind the to-be-remembered and the to-be-forgotten subset to separate group cues.

Method

Participants. Twenty-four new students of the University of Zurich took part in two one-hour sessions in exchange for course credit or 30 Swiss Francs.

Materials and Procedure. Words were drawn from the same pool as for Experiment 1. The procedure was identical to Experiment 1, with the following two changes. First, rather than presenting the words across two rows of frames, the six frames were arranged on a circle, thereby removing any spatial separation between subsets. Words were presented in clockwise order, starting at 10 o'clock; without additional pause between the first and second subset. Second, in the set-size 3 and the set-size 6-3 conditions, on each trial a random subset of 3 words was selected as relevant. These words were cued by turning their frames red 1s after offset of the last word. In the set-size 3 condition, the three irrelevant words were replaced by "XXXX".

Results and Discussion

Mean RT and proportion correct are presented in Figure 4. The BFs for all effects of interest are given in Table 1. There is no hint of removal of the to-be-forgotten subset in condition 6-3: Regardless of CTI, both RTs and accuracies remained at the level of set-size 6. This result confirms that, when the to-be-forgotten subset is defined *ad hoc* after encoding of the entire memory set, the set of items competing for retrieval cannot be reduced.

Experiment 3

With Experiment 3 I investigated whether item-context bindings can be removed if the relevance cue is provided immediately after encoding each item, and if so, how long removal takes. Based on the value of the removal-rate parameter in the SOB-CS model [27], I expected removal to take about 1.5s to be completed.

Method

Participants. Twenty-four new students of the University of Zurich took part in two one-hour sessions in exchange for course credit or 30 Swiss Francs.

Materials and Procedure. Materials and procedure were the same as in Experiment 2, with the following changes: Immediately after offset of each word, the frame of that word turned green for a to-be-remembered word, or turned red for a to-be-forgotten word. If the word was to be remembered, the next word was presented simultaneously with the remember cue for the preceding word. If the word was to be forgotten (or was replaced by "XXXX" in the set-size 3 condition), the next word was presented after a variable interval, set to the CTI levels of Experiments 1 and 2, divided by 3. In this way, the time for removing three items in the preceding experiments (i.e., the CTI) was divided equally between the three to-be-forgotten items in the present experiment. In the set-size 6 condition, a random subset of three to-be-remembered words was followed by the CTI/3 interval, so that the free time interspersed between word presentations was equated for all conditions. Immediately after offset of the last word (if it was to-be-remembered), or after the last removal interval (if it was to-be-forgotten), the recognition probe was presented in one of the frames of relevant words.

Results and Discussion

Mean RTs and proportion correct are shown in Figure 5; BFs can be found in Table 1. Both RT and accuracy data show that the irrelevant words in the 6-3 condition were completely removed already at the shortest CTI levels. The shortest CTI (100 ms) provides a mere 33 ms for removing a to-be-forgotten word before onset of the next word. It is possible that participants used part of the presentation time of the next word (1s) to remove a preceding to-be-forgotten word. Doing so, however, would have curtailed the time for encoding some of the to-be-remembered words. This should have led to poorer memory for these words in the 6-3 condition, compared to the set-size 3 condition in which nothing ever needs to be removed. This was not found, so that any time borrowed from encoding of the next word for removing the preceding one must have been minimal.

We are left with two theoretical options: One is that a just-encoded word can be removed completely from WM – including its activation and its item-context bindings – within a very short time (no more than a few 100 ms). The other is that participants withhold encoding of each word until they see the cue, and then encode only to-be-remembered words into WM. This would require that a representation of the word can be maintained outside WM at least for a few 100 ms after it disappears from the screen, and that this short amount of time is sufficient to encode the word with sufficient strength to enable memory accuracy at about the same level as in the preceding two experiments, in which there was no point in postponing encoding. I ran Experiment 4 to adjudicate between these two options.

Experiment 4

If participants in Experiment 3 postponed encoding of each word until they saw the cue, they must have maintained a representation of that word at least until they saw and processed the cue, and encoded the word into WM. Assuming that encoding takes not much more than 500 ms, it is conceivable that the word is maintained in sensory memory. If that is the case, the postponement strategy would break down if the cue is delayed for 1 s after offset of each word. This is what I did in Experiment 4.

Method

Participants. Twenty-four new students of the University of Zurich took part in two one-hour sessions in exchange for course credit or 30 Swiss Francs.

Materials and Procedure. Materials and procedure were the same as in Experiment 3, with a single difference: I inserted a 1s interval between offset of each word and onset of the cue informing participants whether to remember (green frame) or to forget (red frame) that word.

Results and Discussion

Figure 6 and Table 1 present the results. Removal of the irrelevant words in condition 6-3 was just as instantaneous and complete as in Experiment 3. This result renders it very unlikely that participants did not encode each word into WM until they saw the cue. To do so they would have found a way to remember that word for at least 1s – more than is assumed to be possible using visual sensory memory. We are forced to conclude that each word is encoded into WM before the onset of the cue. Hence, the to-be-forgotten words must have been removed from WM rapidly and completely.

One possibility is that each word, upon presentation, is maintained in the focus of attention of WM [42] without being bound to its context. In SOB-CS, this could be implemented by keeping the representation of the just-seen word, together with its position, active for a while without applying the Hebbian learning process. In Figure 3, this would be a state in which a single item and a single position node are active, but not yet bound to each other. As long as only a single word and a single position is maintained active in this way, this involves no risk of confusing or mixing multiple words or positions with each other. Therefore, when a remember cue is presented 1s later, the currently active word and the currently active position context can still be bound uniquely to each other.

The results of Experiments 3 and 4 are strikingly different from what I predicted from the SOB-CS model: In SOB-CS, representations of memory items and distractors are immediately encoded by binding them to their context, and removal of distractors – by unbinding them from their context – proceeds gradually over the course of 1-2s. The present experiments imply that information not known yet to be relevant is either not bound to a context in WM immediately, or else, removed very rapidly and completely.

These results are also at odds with other evidence speaking to the removal of distractors from WM: In the complex-span paradigms, distractors tend to be erroneously recalled instead of memory items [43]. This is found even after fairly long periods of free time following each distractor, during which they could have been removed from WM [30, 44]. These distractor intrusions suggest that removal of distractors is slow, and not complete even after several seconds. One difference between the complex-span paradigm and the present experiments is that in complex span, distractors have to be processed (e.g., reading distractor words or sentences, solving equations). It could be that by being processed, stimuli are encoded into WM in a way that makes them more difficult to remove afterwards, compared to stimuli that are merely attended and briefly maintained – perhaps in the focus of attention – for a potential later memory test. Experiment 5 serves to explore this possibility.

Experiment 5

Experiment 5 was similar to Experiments 3 and 4, except that each word had to be processed in addition to being remembered (or not) for a subsequent memory test. The processing task required a semantic judgment on each word. The cue informing whether or not to remember each word was given immediately after the judgment was made for that word.

Method

Participants. Twenty-four new students of the University of Zurich took part in two one-hour sessions in exchange for course credit or 30 Swiss Francs.

Materials and Procedure. Memory items were drawn from a new set of German words, all referring to concrete objects. The procedure was the same as in Experiments 3 and 4, with the following differences: Participants were instructed that, upon appearance of each word, they were to judge whether the object the word referred to was typically larger or smaller than a soccer ball. They made their judgment by pressing either the left or the right arrow key. As soon as they pressed an arrow key, or a maximum of 1.8s had elapsed, the word was erased, and at the same time, the frame turned green (for to-be-remembered words) or red (for to-be-forgotten words). Obviously, when a row of X was presented instead of a word in the set-size 3 condition, no judgment was expected, and the stimulus was presented for the full 1.8s, followed by its frame turning red. A red frame (in the set-size 3 and 6-3 conditions) or a random subset of green frames (in the set-size 6 condition) was followed by a variable free time (CTI/3) before onset of the next word. In those cases in which no free time was added, the next word was displayed simultaneously with the previous frame turning green.

Results and Discussion

As can be seen in Figure 7 and Table 1, the results differ markedly from those of Experiments 3 and 4. The accuracy data show no sign of removal; the RT data show evidence for partial removal: RTs in the 6-3 condition were halfway between those in the two baseline conditions. The figure suggests that the RTs in the 6-3 condition moved towards the longer RTs of set-size 6 with increasing CTI, but the statistical evaluation provided evidence against the interaction of CTI and set size (BF = .12 and .05, implying BF = 8 and 20, respectively, in favor of the Null hypothesis).

These findings, together with those of Experiments 3 and 4, suggest that a simple processing operation – making a judgment about a word's meaning – entails encoding of that word into WM in a way that makes it very hard to remove it again right after processing. This is partly in line with findings from complex-span experiments: Distractors, after having been processed, can be removed from WM only partially during a subsequent interval of free time. At the same time, the present findings are at odds with another well-replicated observation with complex-span tasks: As the free time following each distractor is increased, memory performance usually gets better. This finding, first observed by Barrouillet and his colleagues, is known as the *cognitive-load* effect [45, 46]. In the present experiment, the effect of cognitive load should have manifested itself as a beneficial effect of longer CTIs in the 6-3 condition. There was no hint of such an effect in the data. My final experiment follows up on this unexpected absence of the cognitive-load effect by moving the present paradigm closer to the typical complex-span paradigm.

Experiment 6

One difference between the present experiments and typical complex-span experiments is that here, I tested memory through local recognition, whereas in complex span, memory is usually tested through recall. Although there is no theoretical reason to expect the cognitive-load effect to depend on testing method, I found this question worth exploring. Therefore, with Experiment 6 I repeated Experiment 5, using a recall test.

Method

Participants. Twenty-four new students of the University of Zurich took part in two one-hour sessions in exchange for course credit or 30 Swiss Francs.

Materials and Procedure. The experiment was identical to Experiment 5 except for the memory test: Participants were asked to recall three words of the relevant set. In the set-size 6 condition, where the relevant set comprises all six words, a random subset of three was tested on each trial. The three relevant words were probed for recall in a random order by placing a question mark in their frame. Participants had to type at least the first three letters of the word, followed by "Enter".

Results and Discussion

The accuracy data are plotted in Figure 8, and the corresponding BFs are given in Table 1. I did not analyze RTs because the timing of typed responses is probably not a good measure of retrieval speed. There was again evidence for partial removal in the 6-3 condition, this time reflected in the accuracy data. Different from Experiment 5, accuracy in this condition improved with longer free time (BF for the main effect of CTI in condition 6-3 = 50). Although longer free time also appeared to slightly benefit the other two conditions, this effect was not supported statistically (for set-size 3: BF = 0.05; for set-size 6: BF = 0.1). At the same time, we cannot be confident that the effect of CTI was larger in the 6-3 condition than in the two baseline conditions: The evidence goes against the interaction of CTI with set size (6 vs. 6-3, BF = 0.1); the interaction of CTI with set size (3 vs. 6-3) was supported statistically, but might reflect a ceiling effect on set-size 3. The safe conclusion is that there was a beneficial effect of free time, but it is uncertain whether this effect was limited to the 6-3 condition, in which this time could have been used to remove irrelevant words, or whether it also applied to the two baseline conditions in which there was nothing to remove.

Why was the effect of cognitive load observed with the recall test of Experiment 6, but not with the local-recognition test of Experiment 5? Perhaps the absence of the effect in Experiment 5 is no more than a random outlier in a long series of successful replications of the cognitive-load effect; before we draw strong conclusions, this unexpected finding should be replicated. Assuming that it is real, one possibility is that the cognitive-load effect emerges only when participants need to reproduce the memory content. This is unlikely, because a beneficial effect of free time was also found in three experiments in which participants selected the correct word from a matrix of 15 or 18 alternatives [30]. Another possibility is that there is no cognitive-load effect for the first item tested. In Experiment 5 only one item was tested, whereas in Experiment 6, three items were tested. This, too, can be ruled out because the effect of free time in Experiment 6 was strongly supported for condition 6-3 when the analysis is limited to the first word tested (BF = 38). The remaining difference is that Experiment 5 required a yes-no decision on a test probe, whereas all experiments so far that showed the cognitive-load effect required reproduction or selection of the memory items. I cannot think of a theoretical reason why the test format per se should determine whether free time can be used for removing irrelevant material from WM (or for any other beneficial process). One possibility is that participants in Experiment 5 could have made use of the free time, but because they experienced the memory test as relatively easy, they chose not to.

General Discussion

The present exploration of the removal of irrelevant information from WM revealed an astonishing diversity of findings: When an individual stimulus is cued to be irrelevant shortly after encoding, it can be removed instantaneously and completely (item-wise removal; Experiments 3 and 4) – unless the stimulus has been processed before, in which case removal remains incomplete (Experiments 5 and 6). When a subset of items is cued to be irrelevant after the entire memory set has been encoded (set-wise removal), that subset can be removed from WM only if it has been marked as a separate group already at encoding (Experiments 1 and 2).

In addition to the variety of time courses of removal – from extremely fast to not at all – the present experiments necessitate a distinction between two kinds of removal: Item-wise removal affects both RT and accuracy, whereas set-wise removal affects only RT. I propose that set-wise removal removes the irrelevant items only from the set of representations competing for retrieval, but not from the set competing for the capacity limit of WM. In contrast, item-wise removal also eliminates an item from the competition for limited capacity, so that the remaining items can be remembered as well as if the irrelevant item had never been encoded.

The results can be understood as follows: Immediately after encoding, the last-encoded stimulus and its context can be maintained in an active state. I define this *active state* in terms of connectionist network models, such as SOB-CS, in which a distinction is made between the current pattern of activation across a set of units, constituting the currently active representation(s), and the pattern of connection weights between units, implementing bindings between elementary representations. This distinction can be tentatively mapped onto the distinction between "neurally active" and "neurally silent" representations in WM that is emerging in cognitive neuroscience [47-49]. Representations that are active simultaneously can be bound together through Hebbian learning, or unbound through Hebbian anti-learning. I will refer to representations in an active state as being in the *focus of attention* in WM. As long as the last-encoded stimulus is still in an active state, it can be removed virtually instantaneously. This could be so for two reasons: (1) The stimulus is immediately bound to its context; when it is cued as irrelevant, it is rapidly unbound through anti-learning. (2) The stimulus is only maintained in the focus of attention; only once it is cued as relevant, it is rapidly bound to its context. At present I see no compelling reason to decide for one or the other theoretical option. The speed at which a stimulus apparently can be removed (option 1) or encoded (option 2) was unanticipated from the perspective of SOB-CS, in which these processes are assumed to occur more gradually. It converges, however, with the recent observation of very rapid removal in another series of experiments [50].

When the stimulus needs to be processed, at least one new active representation is being generated. In the case of a semantic judgment, this could be a representation of the word's meaning (in addition to the initial visual and phonological code), and a representation of the decision (e.g., that the object is larger than a soccer ball). These new representations tend to replace the initial active representation of the word from the focus of attention. Hence, to maintain the word in memory, its initial representation has to be bound to its context before being replaced. After being replaced, because it is no longer active, it cannot be unbound through anti-learning. Only the products of the processing operation that are still active in the focus of attention can be unbound. Therefore, complete removal after a processing operation is difficult. The more complex the processing operation – the more new representations it creates – the harder it should be to remove them all.

Removing a subset of items from WM after encoding the entire set is difficult for similar reasons: To unbind earlier-encoded items from their contexts, each of these items needs to be brought back into an active state one by one, together with their contexts (i.e., their list positions), to apply Hebbian anti-learning to them. Doing so comes down to retrieving each to-be-removed item. Because this is an error-prone process, especially when WM load is high, the WM system does not usually attempt to unbind these items.

When the subset cued to be irrelevant forms a group already established at encoding, then this group is bound to a group context by which it is distinguished from the other, relevant items. This group context can be used to limit the set of competitors for retrieval in two ways. One is to selectively activate the relevant group context, and through it, the relevant items bound to it (as illustrated in Figure 3B). Alternatively, the irrelevant group context could be de-activated below baseline, thereby de-activating the items bound to it, effectively removing them from the set of competitors.

De-activation below baseline is the definition of *inhibition*, which has been argued to contribute to the control of WM contents [6]. In the theory sketched above, inhibition could fulfil one function in the removal mechanism – reducing the set of retrieval competitors –, and it is complemented by another process, unbinding of contents from their contexts. Inhibition on its own can eliminate the effects of irrelevant WM contents on retrieval speed, but not on accuracy. Unbinding, in contrast, reduces interference between content-context bindings in WM, thereby reversing the effect of irrelevant WM contents on accuracy.

The notion that multiple items can be simultaneously active, thereby forming the set of items competing for retrieval, implies that multiple items are held in the focus of attention simultaneously. I have argued elsewhere that the focus of attention in WM usually narrows down to one item, because multiple simultaneously active items would be superimposed, creating an inextricable blend of several contents that is often dysfunctional for processing the information [42, 51]. In light of the present findings, and findings from neuroscience methods tracking neurally active representations in WM [37, 52], I came to the conclusion that it is possible to maintain multiple items active in the focus of attention during the retention interval. A blend of several active representations is not dysfunctional as long as there is no need to access an individual item within the set (e.g., for updating it, manipulating it, or reporting it). Because each item, in addition to being active, is also maintained in WM through its binding to a unique context (i.e., the position context), it is always possible to narrow down the set of active items to one when necessary – this requires an activation boost specifically to the selected item through its position context, together with de-activation of the group context (see Figure 3C). Keeping multiple items simultaneously active in the focus of attention in preparation for a memory test might actually be functional because the active set defines a small set of competitors for retrieval. The more this set can be constrained ahead of time, the faster a single item within that set can be retrieved. This is reflected in faster access to an item from that set, but not necessarily in higher accuracy, as in Experiment 1 and previous similar observations [21, 37, 38].

References

1. Cowan, N., *Working memory capacity*. 2005, New York: Psychology Press.
2. Oberauer, K., *Design for a working memory.* Psychology of Learning and Motivation: Advances in Research and Theory, Vol 51, 2009. **51**: p. 45-100.
3. Carver, R.P., *Reading rate: Theory, research, and practical implications.* Journal of Reading, 1992. **36**: p. 84-95.
4. Henderson, J.M. and A. Hollingworth, *High-level scene perception.* Annual Review of Psychology, 1999. **50**: p. 243-271.
5. Bjork, R.A., *Positive forgetting: The noninterference of items intentionally forgotten.* Journal of Verbal Learning and Verbal Behavior, 1970. **9**: p. 255-268.
6. Hasher, L., R.T. Zacks, and C.P. May, *Inhibitory control, circadian arousal, and age*, in *Attention and Performance*, D. Gopher and A. Koriat, Editors. 1999, MIT Press: Cambridge, MA. p. 653-675.
7. Altmann, E.M., *Evidence for temporal decay in short-term episodic memory.* Trends in Cognitive Sciences, 2009. **13**: p. 279.
8. Oberauer, K. and S. Lewandowsky, *Further evidence against decay in working memory.* Journal of Memory and Language, 2014. **73**: p. 15-30.
9. Oberauer, K. and S. Lewandowsky, *Evidence against decay in verbal working memory.* Journal of Experimental Psychology: General, 2013. **142**: p. 380-411.
10. Lewandowsky, S. and K. Oberauer, *No evidence for temporal decay in working memory.* Journal of Experimental Psychology-Learning Memory and Cognition, 2009. **35**(6): p. 1545-1551.
11. Berman, M.G., J. Jonides, and R.L. Lewis, *In search of decay in verbal short-term memory.* Journal of Experimental Psychology: Learning, Memory, and Cognition, 2009. **35**: p. 317-333.
12. Ricker, T.J. and N. Cowan, *Loss of visual working memory within seconds: The combined use of refreshable and non-refreshable features.* Journal of Experimental Psychology: Learning, Memory, and Cognition, 2010. **36**: p. 1355-1368.
13. Ricker, T.J., L.R. Spiegel, and N. Cowan, *Time-based loss in visual short-term memory is from trace decay, not temporal distinctiveness.* Journal of Experimental Psychology: Learning, Memory, and Cognition, 2014. **40**: p. 1510-1523.
14. Lilienthal, L., S. Hale, and J. Myerson, *The effects of environmental support and secondary tasks on visuospatial working memory.* Memory & Cognition, 2014. **42**: p. 1118-1129.
15. Pertzov, Y., et al., *Rapid forgetting prevented by retrospective attention cues.* Journal of Experimental Psychology: Human Learning and Memory, 2013. **29**: p. 1224-1231.
16. Cansino, S., et al., *Effects of aging on interference control in selective attention and working memory.* Memory & Cognition, 2011. **39**: p. 1409-1422.
17. Ecker, U.K.H., K. Oberauer, and S. Lewandowsky, *Working memory updating involves item-specific removal.* Journal of Memory & Language, 2014. **74**: p. 1-15.
18. Ecker, U.K.H., S. Lewandowsky, and K. Oberauer, *Removal of information from working memory: A specific updating process.* Journal of Memory and Language, 2014. **74**: p. 77-90.
19. Kessler, Y. and K. Oberauer, *Working memory updating latency reflects the cost of switching between maintenance and updating modes of operation.* Journal of Experimental Psychology: Learning, Memory, and Cognition, 2014. **40**: p. 738-754.
20. Kessler, Y. and N. Meiran, *Two dissociable updating processes in working memory.* Journal of Experimental Psychology: Learning, Memory, and Cognition, 2008. **34**: p. 1339-1348.
21. Oberauer, K., *Removing irrelevant information from working memory: A cognitive aging study with the modified Sternberg task.* Journal of Experimental Psychology-Learning Memory and Cognition, 2001. **27**(4): p. 948-957.

22. Landman, R., H. Spekreijse, and V.A.F. Lamme, *Large capacity storage of integrated objects before change blindness.* Vision Research, 2003. **43**(149-164).
23. Griffin, I.C. and A.C. Nobre, *Orienting attention to locations in internal representations.* Journal of Cognitive Neuroscience, 2003. **15**: p. 1176-1194.
24. Souza, A.S., L. Rerko, and K. Oberauer, *Unloading and reloading working memory: Attending to one item frees capacity.* Journal of Experimental Psychology: Human Perception and Performance, 2014. **40**: p. 1237-1256.
25. Williams, M., et al., *The benefit of forgetting.* Psychonomic Bulletin & Review, 2013. **20**: p. 348-355.
26. Muter, P., *Very rapid forgetting.* Memory & Cognition, 1980. **8**: p. 174-179.
27. Oberauer, K., et al., *Modeling working memory: An interference model of complex span.* Psychonomic Bulletin & Review, 2012. **19**: p. 779-819.
28. Daneman, M. and P.A. Carpenter, *Individual differences in working memory and reading.* Journal of Verbal Learning and Verbal Behavior, 1980. **19**: p. 450-466.
29. Turner, M.L. and R.W. Engle, *Is working memory capacity task dependent?* Journal of Memory and Language, 1989. **28**: p. 127-154.
30. Oberauer, K. and S. Lewandowsky, *Control of information in working memory: Encoding and removal of distractors in the complex-span paradigm.* Cognition, 2016. **156**: p. 106-128.
31. Ryan, J., *Temporal grouping, rehearsal and short-term memory.* Quarterly Journal of Experimental Psychology, 1969. **21**: p. 148-155.
32. Oberauer, K., *Binding and inhibition in working memory: Individual and age differences in short-term recognition.* Journal of Experimental Psychology-General, 2005. **134**(3): p. 368-387.
33. Thalmann, M., M. Niklaus, and K. Oberauer *Estimating Bayes Factors for linear models with random slopes on continuous predictors*. PsyArXiv, 2017. DOI: http://doi.org/10.17605/OSF.IO/7ZPK3.
34. R_Core_Team, *R: A language and environment for statistical computing*. 2017, R Foundation for Statistical Computing: Vienna, Austria.
35. Rouder, J.N., et al., *Default Bayes factors for ANOVA designs.* Journal of Mathematical Psychology, 2012. **56**: p. 356-374.
36. Wagenmakers, E.-J., et al., *Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method.* Cognitive Psychology, 2010. **60**: p. 158-189.
37. LaRocque, J.J., et al., *Decoding attended information in short-term memory: an EEG study.* Journal of Cognitive Neuroscience, 2013. **25**(1): p. 127-42.
38. Shepherdson, P., K. Oberauer, and A.S. Souza, *Working memory load and the retro-cue effect: a diffusion model account.* Journal of Experimental Psychology: Human Perception and Performance, 2017.
39. Oberauer, K., *What is working memory capacity? / ¿Qué es la capacidad de la memoria de trabajo?* Estudios de Psicología / Studies in Psychology, 2017. **38**: p. 338-384.
40. Farrell, S., *Temporal clustering and sequencing in short-term memory and episodic memory.* Psychological Review, 2012. **119**: p. 223-271.
41. Anderson, J.R. and M. Matessa, *A production system theory of serial memory.* Psychological Review, 1997. **104**: p. 728-748.
42. Oberauer, K. and L. Hein, *Attention to information in working memory.* Current Directions in Psychological Science, 2012. **21**: p. 164-169.
43. Oberauer, K., et al., *Interference between maintenance and processing in working memory: The effect of item-distractor similarity in complex span* Journal of Experimental Psychology: Learning, Memory, and Cognition, 2012. **38**: p. 665-685.
44. Hoareau, V., et al., *Computational and behavioral investigations of the SOB-CS removal mechanism in working memory*, in *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* G. Gunzelmann, et al., Editors. 2017, Cognitive Science Society: London. p. 532-537.

45. Barrouillet, P., et al., *Time and cognitive load in working memory.* Journal of Experimental Psychology: Learning, Memory & Cognition, 2007. **33**: p. 570-585.
46. Barrouillet, P., S. Bernardin, and V. Camos, *Time constraints and resource sharing in adults' working memory spans.* Journal of Experimental Psychology: General, 2004. **133**: p. 83-100.
47. Postle, B.R., *How does the brain keep information "in mind"?* Current Directions in Psychological Science, 2015. **25**: p. 151-156.
48. Stokes, M.G., *'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework.* Trends in Cognitive Sciences, 2015. **19**(7): p. 394-405.
49. Sprague, T.C., E.F. Ester, and J.T. Serences, *Restoring latent visual working memory representations in human cortex.* Neuron, 2016. **91**: p. 694-707.
50. Oberauer, K. and S. Lewandowsky, *Simple measurement models for complex working memory tasks*. 2017: Unpublished Manuscript, University of Zurich.
51. Oberauer, K. *The focus of attention in working memory - from metaphors to mechanisms*. frontiers in Human Neuroscience, 2013. **7**, DOI: 10.3389/fnhum.2013.00673.
52. Lewis-Peacock, J.A., et al., *Neural evidence for a distinction between short-term memory and the focus of attention.* Journal of Cognitive Neuroscience, 2012. **24**(1): p. 61-79.
53. Bakeman, R. and D. McArthur, *Picturing repeated measures: Comments on Loftus, Morrison, and others.* Behavioral Research Methods, Instruments, & Computers, 1996. **28**: p. 584-589.

Table 1: Bayes Factors in favor of each effect in the BANOVAs of Experiments 1 to 5.

Effect	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6
Log(RT), Set Size 6-3 vs. Set Size 6						
Set Size	8.0×10^{7}	0.04	2.8×10^{27}	1.6×10^{35}	28207	
CTI	0.11	0.09	0.07	0.8	0.09	
Set Size x CTI	2.1×10^{4}	30	0.07	0.04	0.12	
Log(RT), Set Size 6-3 vs. Set Size 3						
Set Size	6.7×10^{3}	1.3×10^{32}	0.07	0.05	2.4	
CTI	0.08	0.12	0.15	1.3	0.9	
Set Size x CTI	6.4×10^{7}	9	1.7	0.04	0.05	
Proportion Correct, Set Size 6-3 vs. Set Size 6						
Set Size	0.27	0.05	9.1×10^{3}	17	0.076	1.4×10^{26}
CTI	0.09	0.03	0.15	0.04	0.05	24
Set Size x CTI	0.07	0.05	0.17	0.04	0.06	0.1
Proportion Correct, Set Size 6-3 vs. Set Size 3						
Set Size	263	4.9×10^{9}	0.07	0.16	2.3×10^{4}	2.4×10^{3}
CTI	0.05	0.04	0.03	0.03	0.04	7.1
Set Size x CTI	0.05	0.05	0.04	0.06	0.06	28

Figure Captions

Figure 1. Sequence of events in a trial of Experiment 1 (left) and Experiment 2 (right), illustrating the 6-3 condition. In Experiment 1, six words are presented sequentially across two rows of frames. One of the rows is marked as relevant by its frames turning red. In Experiment 2, the six words are presented in clockwise order across a circular arrangement of frames. A random set of three frames is marked as relevant. In both experiments, one frame selected randomly from the relevant subset is tested by a probe presented in that frame. This happens aver a variable cue-target interval (CTI).

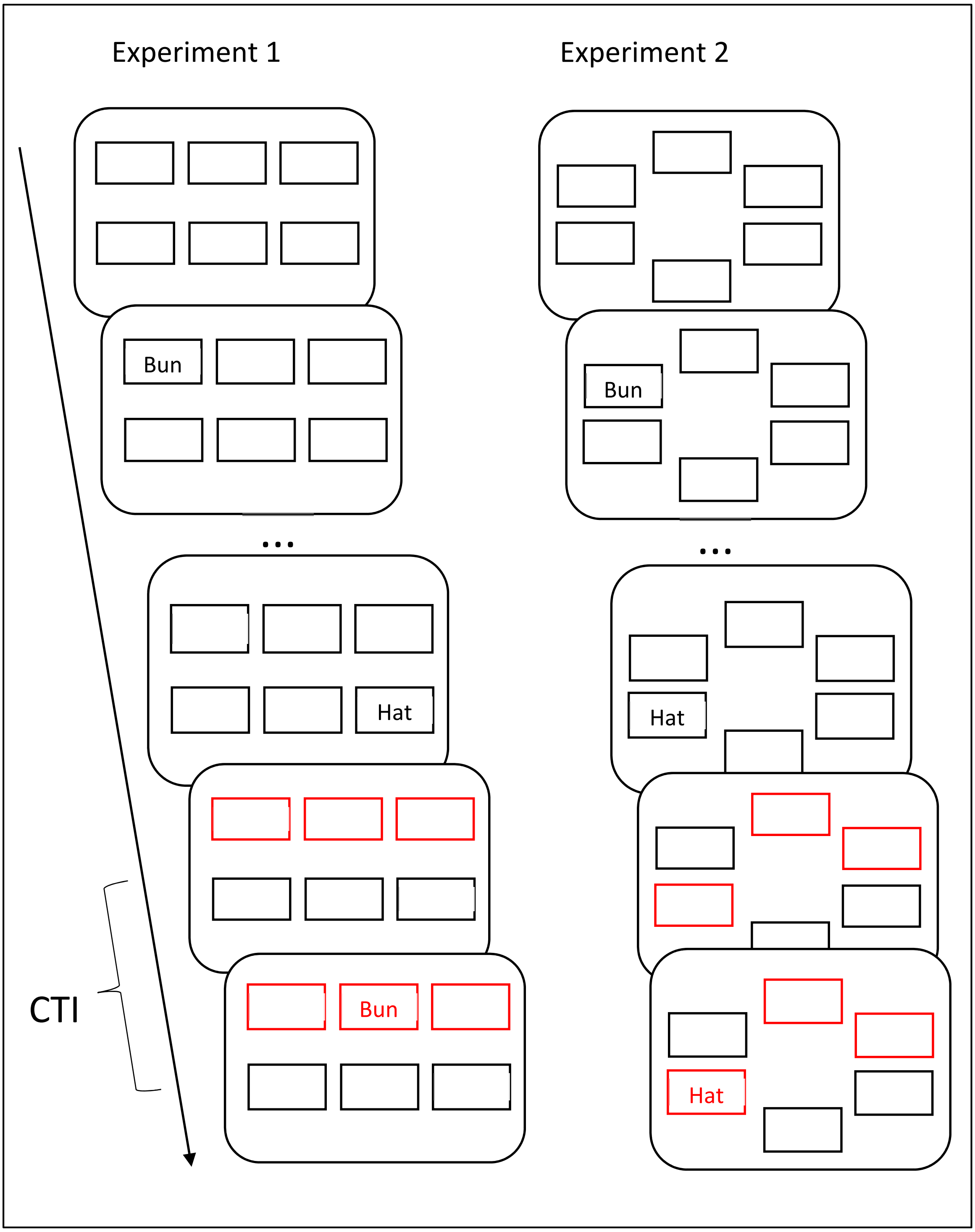

Figure 2. Mean response times (left) and accuracies (right) in Experiment 1. Error bars are 95% confidence intervals for within-subjects comparisons [53].

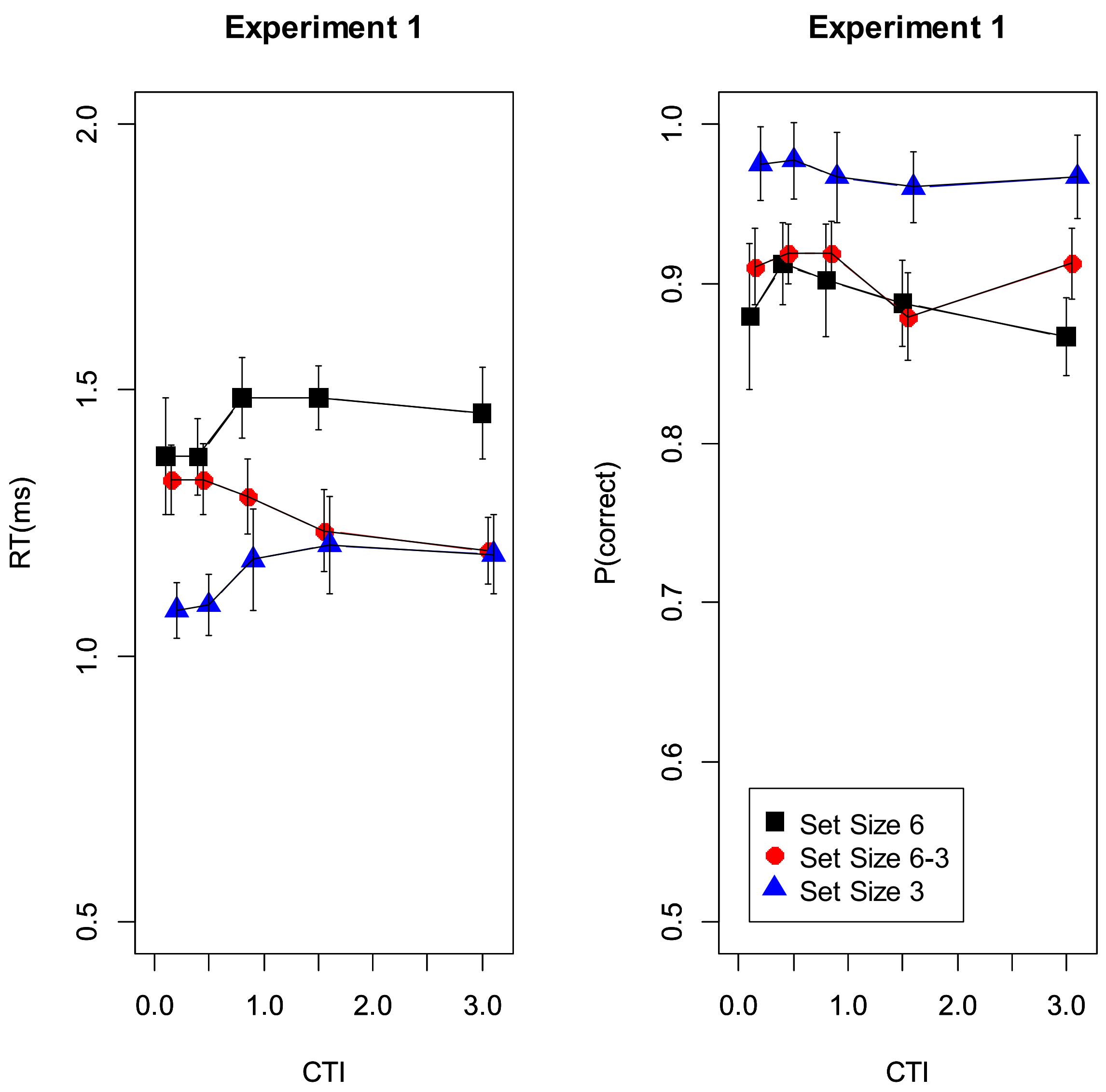

Figure 3: Representational structure in WM for a memory set subdivided into two subsets (groups) at encoding. A: Immediately after encoding, both group contexts are active, activating all memory items. B: After the retro-cue to the top group, only the top-group context remains active, so that only the three items in the top remain active. C: After the recognition probe in one position, the corresponding position context becomes active, conferring activation uniquely to the target item bound to that position; the relevant group's context is de-activated to reduce competition from the other items in that group.

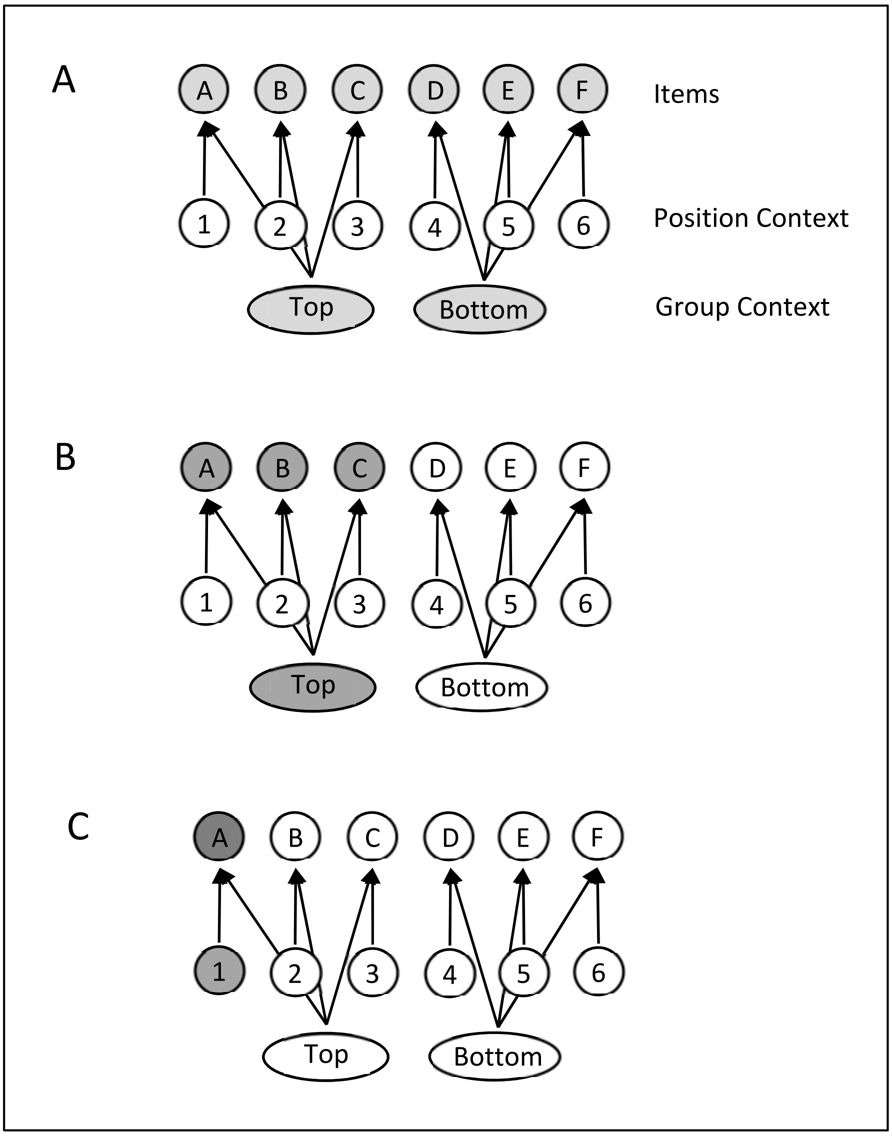

Figure 4. Mean response times (left) and accuracies (right) in Experiment 2.

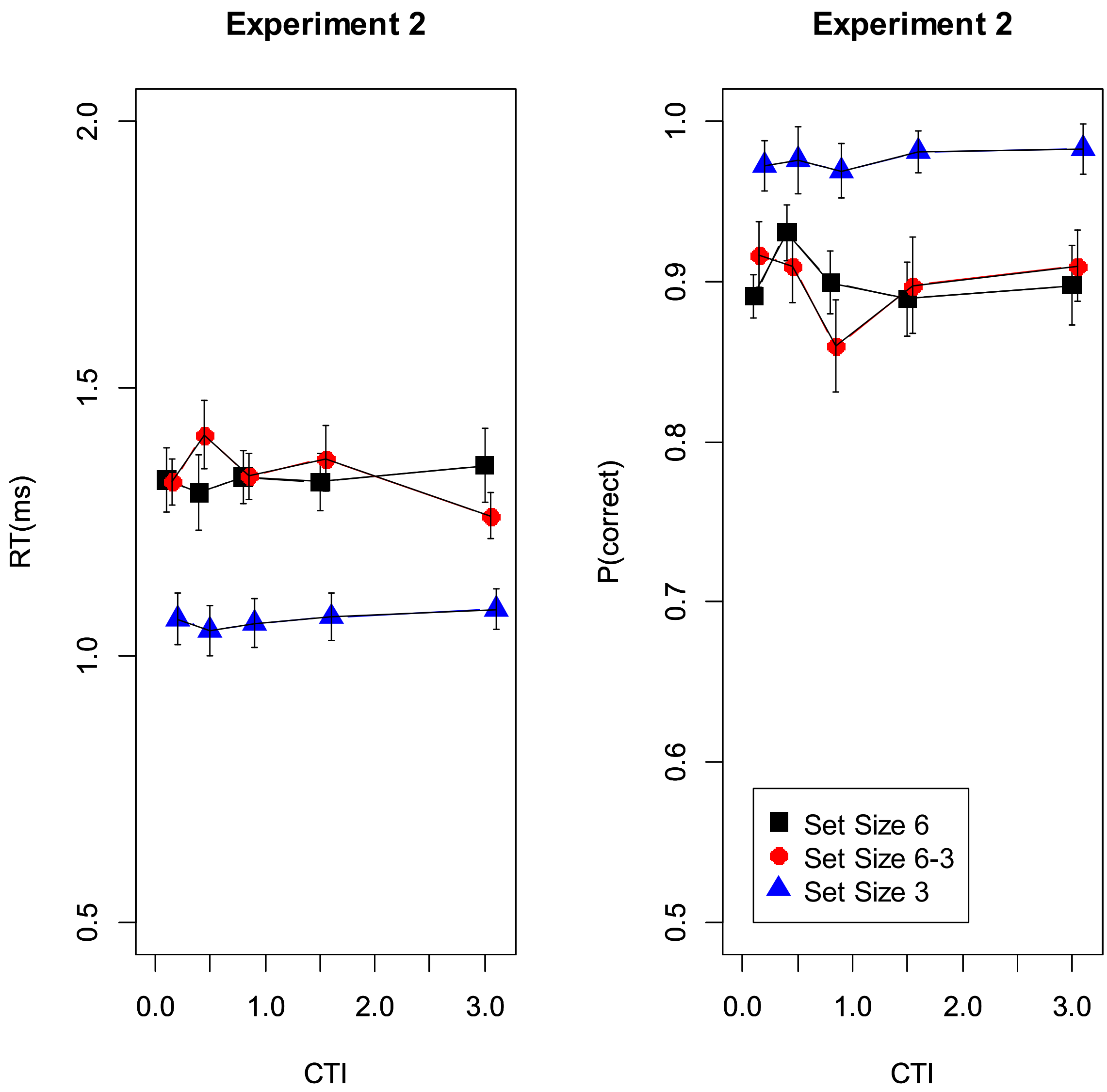

Figure 5. Mean response times (left) and accuracies (right) in Experiment 3.

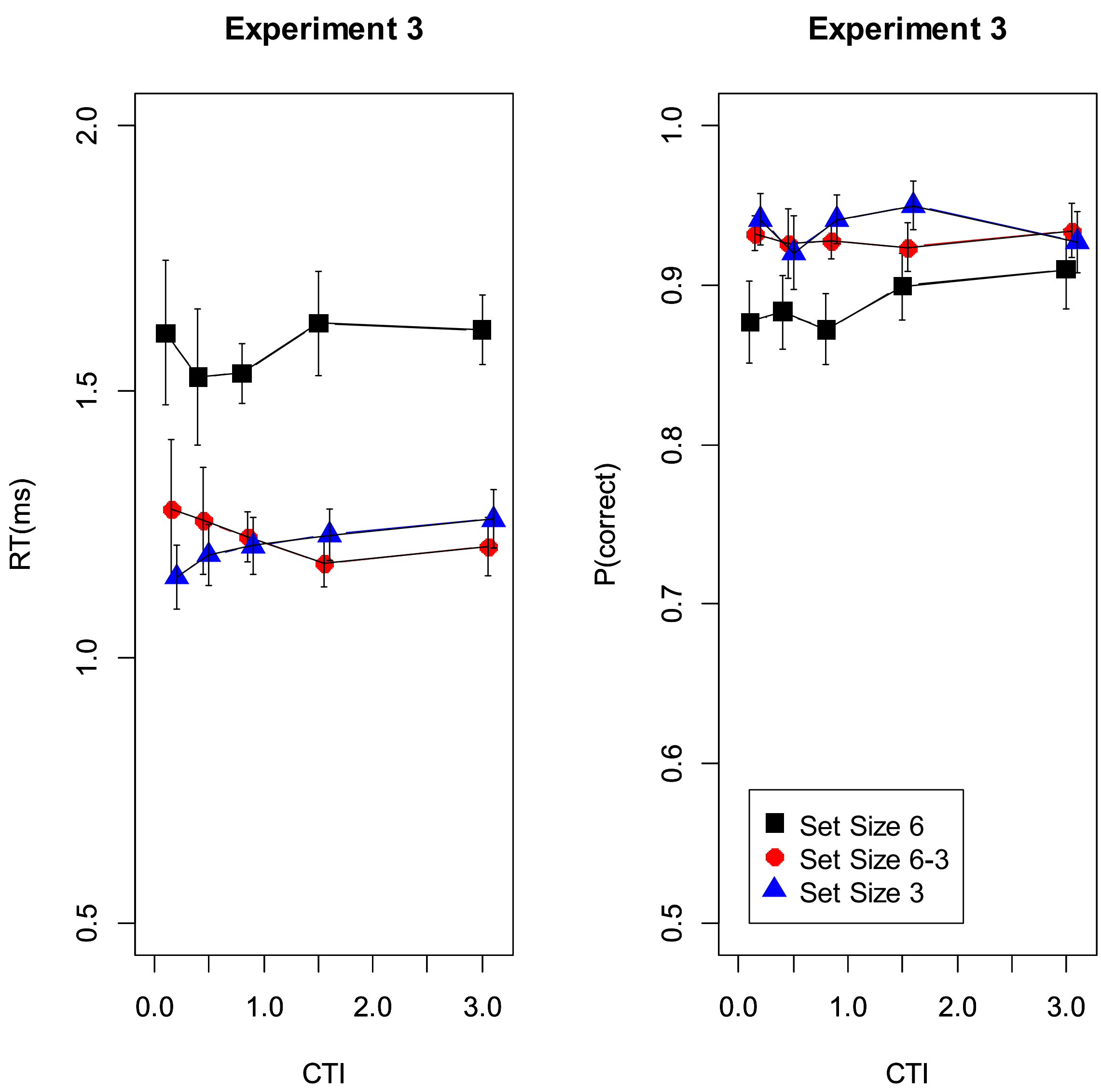

Figure 6. Mean response times (left) and accuracies (right) in Experiment 4.

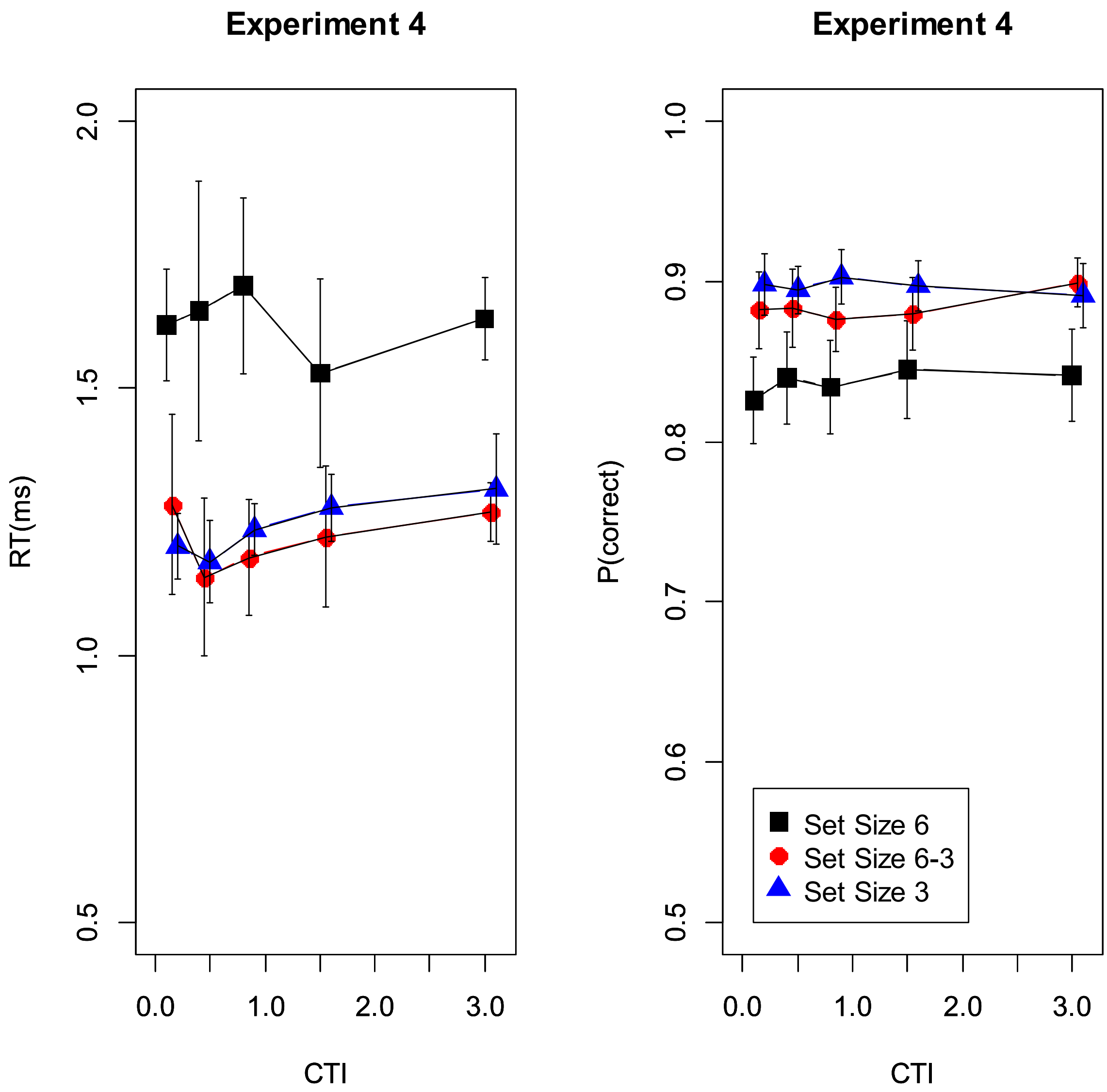

Figure 7. Mean response times (left) and accuracies (right) in Experiment 5.

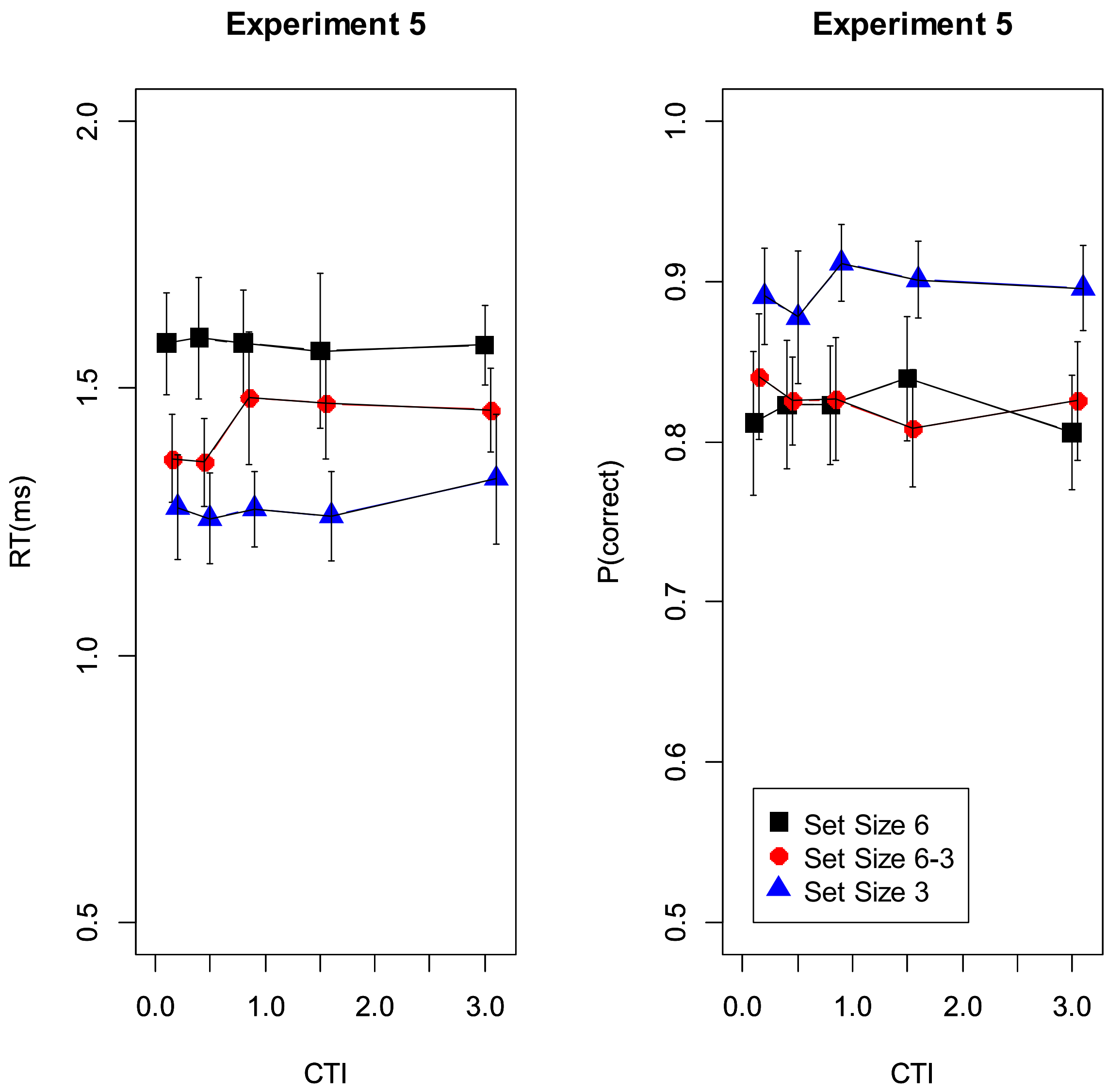

Figure 8. Mean recall accuracies in Experiment 6.